UNIVERSITY OF ILLINOIS

April 17 19 89

THIS IS TO CERTIFY THAT THE THESIS PREPARED UNDER MY SUPERVISION BY

Sharon L. Frey

ENTITLED Network Analysis as Applied to a Group of AIDS

Patients Linked by Sexual Contact

IS APPROVED BY ME AS FULFILLING THIS PART OF THE REQUIREMENTS FOR THE

DEGREE OF Bachelor of Science in Psychology

Stanley Wasserman

Instructor in Charge

APPROVED:

HEAD OF DEPARTMENT OF

O-1364

Network Analysis as Applied to a Group of AIDS Patients Linked by Sexual Contact

By

Sharon L. Frey

Thesis

for the
Degree of Bachelor of Science
in
Psychology

College of Liberal Arts and Sciences
University of Illinois
Urbana, Illinois

1989

Table of Contents

Abstract

Numerous reports of two rare diseases, Kaposi's sarcoma and Pneumocystis carnii pneumonia, were received by the Centers for Disease Control in 1981. General practitioners in Los Angeles, New York City, and San Francisco reported the incidence of these diseases in homosexual men who had no underlying symptomotology for immunological disease. Studies of these cases suggested a common link between the occurrence of the disease and an unexplained cell-mediated immunodeficiency. This particular immunodeficiency has since been termed the "Acquired Immune Deficiency Syndrome", or as it is better known, AIDS.

Klovdahl (1985) states that a "network approach", providing information in graphic form and showing the spread of an infectious agent transmitted from one individual to another, has been implicit in the study of epidemics (epidemiology) for many years. Klovdahl contends that conceptualizing a population as a social network can aid in understanding the nature and spread of disease, by viewing individuals not merely as randomly interacting actors, but as linked or connected through personal contact. Knowledge of a network's structure then can be used to predict the rate and extent of the spread of disease. Low rates of interaction, as opposed to high rates, lead to diseases which are confined to only a few isolated subpopulations and likely to spread slowly. Networks can also change over time which can

have important implications for the spread of some diseases, and can be investigated using probabilistic models.

Network Analysis departs from traditional, individualistic approaches by recognizing the significance of the social context in which behavior occurs. Two basic underlying assumptions of network analysis are: 1) a person participates in a social system involving other people; and 2) various levels of social structure may exist which may have consequences over and above the effects of the characteristics and behavior of the individuals involved.

Measurements and indices derived from graph theory, a branch of mathematics, are reviewed and applied to a network of linkages among a set of patients with the Acquired Immune Deficiency Syndrome (AIDS), presented by Auerbach, Darrow, Jaffe, and Curran in 1984. Further analysis of the network is accomplished by computer simulation using UCINET, a computer package for the purpose of mathematically studying and manipulating social networks. The intention of these analyses is the investigation of the implications of a network probability model for the spread of AIDS.

Network Analysis as Applied to a Group of AIDS Patients Linked by Sexual Contact

Introduction

A number of diagrams depicting sets of socially interacting individuals has appeared in the medical and epidemiologic literature. These usually describe individuals afflicted with a particular disease and the social contacts that connect them. The diagrams are provided to support the hypothesis of an infectious agent causing a disease of unknown etiology (Klovdahl,1986). Studies have included cases of poliomyelitus (Pearson & Rendtorff, 1944), Hodgkin's disease (Vianna, Greenwald, & Davies, 1971; Vianna, Greenwald, Brady, Polan, Dwork, Mauro & Davies, 1972; Klinger & Minton, 1973), and leukemia and lymphoma patients (Schimpff, Schimpff, Brager, & Weirnik,1975).

The rate and extent of a disease transmitted through personal relations depends on the structure of the network involved. If the network is relatively disconnected, the disease will be confined to a small group of individuals and spread slowly to the rest of the population. If the individuals are closely connected and interaction high, then a more rapid and extensive spread is possible. Therefore, knowledge of a network's structure is helpful in projecting the rate and extent of the spread of infectious diseases, in developing more effective strategies for limiting the effects of some infections, and in anticipating the consequences of identifying different containment

strategies (Klovdahl,1985).

Recently, the Acquired Immune Deficiency Syndrome, AIDS, one of the most publicized and most feared diseases of modern times, has been investigated using a network approach. The Centers for Disease Control define AIDS as "a reliably diagnosed disease that is at least moderately indicative of an underlying cellular immunodeficiency when no known cause or reduced resistance to that disease is present" (Selik, Haverkos, & Curran, 1984). AIDS is caused by a virus formerly called HTLV-III (Human T-cell Leukemia Virus Type III), or LAV (Lympadenopathy Associated Virus), but more commonly referred to as HIV (Human Immunodeficiency Virus). This virus attacks the immune system, leaving the body completely vulnerable to secondary infections which overcome the victim and ultimately lead to death. Among these secondary infections are Kaposi's sarcoma, Pneumocystis carnii pneumonia, tuberculosis, meningitis, toxoplasmosis, mycobacterium, and cytomegalovirus. AIDS, is not confined to one segment of the population, affecting men, women, and children, with homosexual and bisexual men (non-intravenous drug users) accounting for the largest single group affected (66% of all cases in 1987) (Tyckoson, 1988). The only known means by which AIDS is transmitted are through sexual contact, contaminated blood products, shared intravenous drug needles, and in utero (and possibly passed by mother's milk) from mother to child (Paine, 1988).

In one of the earliest studies of AIDS patients, Auerbach, Darrow, Jaffe, and Curran (1984), presented a diagram showing clusters of socially interacting individuals and the linkages among these patients.

Klovdahl (1985) examined the AIDS network provided by Auerbach et al. (1984). Using a computer-generated visual representation, Klovdahl considered the implications of a network probability model for the infectious agent hypothesis. A simple analysis of these data (incorporating three tests: personal contact, temporal sequence, and non-randomness) supported the infectious agent hypothesis and demonstrated that personal contacts (sexual relations) allowed the transmission of this agent between patients with AIDS.

The intention of the present study is to analyze the AIDS network presented by Auerbach et al. as a "typical" AIDS network using statistical models. From this network, other networks can be "created" by manipulating the parameters of the prototype network model. In this way the change in a network's structure, and thus the spread of the disease, can be examined. By examining the structure of an AIDS network, one can determine the most appropriate points of entry for a possible vaccination to prevent further spread. However, one must first understand the nature of networks, how they can be studied, and what information can be gained by their use. This review is by no means comprehensive, but should be viewed as an introduction to some of the concepts involved in network analysis.

What is a Social Network?

A network is a distinct set of entities (persons, objects, or events), termed actors or nodes, that are linked by a specific type of relation or meaningful social tie (Mitchell, 1969). Network analysis uses relational information as its data. Depending on the specific

goals of the individual researcher, networks may vary in both the number of relations examined, from one (unirelational) to many (multirelational), and in the content of these relations: authority and power (Cook & Emerson, 1978); boundary penetration (Mariolis, 1975); communication (Rogers & Kincaid, 1981); instrumental (Granovetter, 1974); kinship and descent (Bott, 1955); sentiment (Hallinan, 1974); and transaction (Burt, Christman & Kilburn, 1980). This focus on relations, characteristics of groups of actors, as opposed to attributes, characteristics of individual actors, is unique to social networks.

Network analysis describes and interprets actor relationships in ways that are radically different from those of conventional behavioral scientists. What network analysts or structuralists propose is the opposite of a priori classification of entities into discrete categories. They propose, first, a set of relations from which the structure of groups can be defined. However, network analysis takes into account not only the relations that exist, but also the relations that do not exist, sometimes termed structural "holes". The configuration of ties between actors exposes a specific network structure. Networks vary greatly in structure, from a set of isolated actors who are not connected to one another, to the complete network in which every actor is connected to every other actor. Social structure is the "regularities in the patterns of relations among concrete entities" (White, Boorman, & Brieger, 1976).

Unlike traditional, individualistic approaches of the social, political, and behavioral sciences, in which the social context in which

actors are involved is generally ignored, network analysis makes two assumptions about social behavior: The first is that an actor typically participates in a social system involving many other actors. The nature of an actor's relationship with others in the social system may affect the actor's perceptions, beliefs, and actions. The second is the importance of the various levels of social structure in a social system. "The structure of relations among actors, and the location of individual actors have important behavioral, perceptual, and attitudinal consequences for the individual actors and for the system of actors as a whole" (Knoke & Kuklinski, 1982).

In analyzing the structural properties of networks in which individual actors are involved and social phenomena can be detected, the organization of social relations becomes central. The principal goal of network analysis is to integrate and develop a set of perspectives and methods that will enable behavioral scientists to investigate more intensely the systematic sources and consequenccs of social behavior in organized settings.

History of the Social Network Paradigm

The "sociometric test" invented by Jacob L. Moreno (1934) as a method for gathering and organizing relational data on the interpersonal attitudes of individuals in small, informal groups is the first application of the network framework. The technique involves asking each member of a group whom in the group they prefer or like. The graphical device Moreno used to represent the data was a "sociogram" in which arrows (relations) linked nodes (actors). Moreno

also introduced terms for specific features of sociometric data, such as "stars" and "isolates." The stars of an network are those individuals who are chosen the most but who do not choose many others in return. Isolates, on the other hand, neither choose nor are chosen by the other actors in the network. Moreno's data representation provided a concrete and appealing image of a social system and created a data structure that could be represented visually and analyzed quantitatively (Leinhardt, 1977).

In the 1950's and early 1960's, graph theory, a branch of mathematics specifying the systematic relations among and between elements (nodes or points) of a graph and the lines (relations or edges) joining them, became incorporated into network analysis as a means of describing the relationships among people, family groups, corporations, government offices, and nation states. There were many advantages to using graph theory for network analysis. It increased researchers' abilility to visualize, calculate, quantify and prove certain theorems.

The first studies using networks dealt with the abstract form of the systems they were modelling instead of either the interactions among their actors -- behavior, or the structure. These early studies dealt strictly with unirelational networks of dyads (pairs of individuals). The strengths of relations as well as the positivity or negativity of ties was not considered.

In the 1960's, two research traditions which had originated in the 1950's merged. One was the analytical and formal mathematical models developed by Anatol Rapoport in which statistical parameters

could be calculated for the probabilities of expected relations in random and biased graphs (Rapoport, 1957). Rapoport's studies allowed researchers to explore the impact of social constraints on patterns of information and contact within specific populations. The other research tradition was the more descriptive and phenomenological approach of anthropologists in mapping patterns of ties among individuals who fell outside the subgroups traditionally identified within populations. Barry Wellman (Wellman, 1983) referred to this reformulation as the social network concept, the idea that social structure is best understood in terms of the interplay between the relations among persons, on the one hand, and the positions and roles that they occupy, on the other. However, it was not until the use of algebraic interpretations and the development of multidimensional scaling that the social network concept became more than merely a useful heuristic (Berkowitz, 1982.)

By the 1970's, network analysis was a distinct paradigm with a theoretical vocabulary and scientific goals. The use of log linear models to summarize and describe categorical data in the form of multiple cross-classifications became increasingly popular (Fienberg & Wasserman, 1981).

The present form of network analysis is far from homogeneous, which is due in part to its diverse origins (epidemiology, anthropology, interorganizational studies, small group research, and political economy) and in part to its relatively rapid development and highly integrative focus.

Graph Theory

Because it is concerned with the characteristics of a set of points and the lines joining them, graph theory is a natural vehicle through which network analysis can express its models of social structure. Graph theory has been utilized to study a multitude of concepts that have been applied to social structures. A few of these include balance theory, originally proposed by Heider (1946) and further developed by Cartwright and Harary (1956), clustering or multiple clique phenomena (Davis, 1969), structural equivalence (Lorrain & White, 1971), transitivity of triads (Holland & Leinhardt, 1977), and models for binary directed graphs (Wasserman, 1978).

Terminology

These diverse applications are all built upon the same basic assumptions of graph theory. A graph, G, is a set of finite points, a_1, a_2, a_3, ...a_n, and a set of lines connecting some pairs of these points (Harary, 1959). The points or nodes are visually represented as dots or circles, while the lines between the points represent relations. Two points, a and b, in graph G are adjacent if there exists a line, ab, between a and b in G. Two lines are adjacent if they share a node in common. A graph is said to be complete if every point is adjacent or connected to every other point in the set. If there exists a one-to-one correspondence between the sets of points in two graphs G_1 and G_2, which preserves their adjacency, the graphs are isomorphic.

Graphs.

An ordinary or *undirected* graph, (Figure1a.) does not take into account the possible directionality of ties -- all ties are assumed to be symmetric or mutual. Relations are reciprocal, if when a chooses b, b also chooses a.

The directed graph or digraph, D, (Figure 1b.) consists of a finite set of points, V, and a collection of ordered pairs of distinct points.

Insert Figure 1 here

Each ordered pair (a, b) is called an arc or directed line. Digraphs allow for symmetric as well as asymmetric ties, when, if a chooses b, b does not choose a. In this case, arrows are used to signify the direction of each relation. An arrow in one direction indicates an asymmetric or unreciprocated tie (Figure 1b., lines ae,ed, dc, bc, and ca), while an arrow between points in both directions represents a mutual or reciprocated tie (Figure1b., lines ab and ec). In both the ordinary graph and the digraph, a pair of points not joined or disconnected is considered a null relation (Figure 1a., absence of line ad, Figure 1b., absence of lines ad, bd, and be).

A signed graph, or s-graph, is a special case of a graph in which relations may be either positive or negative. To distinguish between the types of ties graphically, positive and negative ties are represented as solid and dashed lines, respectively (Figure 2a.). A signed digraph, or s-digraph, is a signed graph and a digraph combined, taking into account not only the sign of the relation but also the

direction (Figure 2b.). Signed graphs are a geometric representation

---

Insert Figure 2 here

---

of binary relations and are used to depict situations in which both a relation and its opposite may occur (Cartwright & Harary, 1956).

Matrix Representation.

For pictorial representation and visual clarity, the configuration of points and connected lines is satisfactory, but in order to perform mathematical analyses, a digraph, D, of n nodes is represented as an (n x n) matrix (Figure 3). By convention, the rows are denoted by ties sent and the columns by ties received. Relations are either present or absent and are enumerated as a "1" or "0" respectively. For any square matrix $A = [a_{ij}]$, the entries appearing on the main diagonal, $a_{11}, a_{22}, \ldots, a_{nn}$, are identity or self-looped relations, which are implicit or explicit ties that a given node "shares" with itself (Berkowitz, 1982). These loops, endomorphisms, can represent real feedback loops for certain relations, such as group relations, but for the most part are seen as implicit. For this reason, the lines in a graph from each node to itself are generally not drawn and the corresponding matrix entries are set to zero.

Insert Figure 3 here

The matrices in Figure 3 correspond to the graphs in Figure 1. The matrix paralleling the ordinary graph (Figure 3a.) is symmetrical, because the ties are mutual or undirected. Node a chooses node b, and b reciprocates, so both a_{12} and a_{21} contain a "1". The digraph matrix in Figure 3b. is not symmetric, due to its directed ties. Nodes receiving asymmetric ties have a "0" in their row but have a "1" in their corresponding columns, or vice versa, indicating the unidirectional relations. Node a chooses node e ($a_{15} = 1$) but node e does not choose node a ($a_{51} = 0$). Null relations result in a set of symmetrically positioned zeros (Figure 3a., $a_{15} = 0$, a_{51}, Figure 3b., $a_{14} = 0$, $a_{41} = 0$).

Measures and Indices

Measures obtained from social networks vary in complexity and range from those concerning only the individual actor to those involving the entire network or multiple networks. Many measures have been developed to give specific information about networks and their structure. Four of the most important of these measures will be described here -- reachability, connectedness, density, and centrality.

For illustration of the measures and indices to be discussed, reference is made to the network presented by Auerbach et al. (1984). The network consists of 40 homosexual men linked by sexual contact and all diagnosed as having AIDS. Thirty-six of the 40 are white (non-hispanic), 3 are hispanic, and one is black. Approximately 57.5% of the cases are diagnosed Kaposi's Sarcoma (KS), 15% as Pneumocystis Carnii Pneumonia (PCP), 22.5% are multiply diagnosed as having both

KS and PCP, and the remaining 5% have some other opportunistic infection. The average age is approximately 36 years.

Originally only 19 patients (residing in the Los Angeles and Orange County area) were included in the study. However, it was found that these 19 were linked to 21 other AIDS patients in other parts of the country (New York, San Francisco, Florida, Georgia, New Jersey, Pennslyvania, and Texas) by previous sexual contact. Information pertaining to the location, diagnosis, and time of onset for each patient are presented in Table 1. Figure 4 illustrates the network graphically,

---

Insert Table 1 here

---

showing the links between these patients as well as diagnosis and location. Figure 5 represents the same network in matrix format with the numbers of the rows corresponding to the actors' numbers in figure 4.

Insert Figures 4 and 5 here

Degree of Point.

The simplest measurements involve only individual actors or points of a graph. The bundle at a point in a graph is the set of all lines incident to that point. The degree of a point is simply the number of these lines. Points in a digraph have both an indegree, which refers to

the number of relations received by a given node (matrix column sums), and an outdegree, which is the number of ties a node directs or sends to other nodes (the matrix row sums). Moreno's "stars", then, have a high indegree (receives many ties) and a low outdegree (sends few ties), while the "isolates" have both a zero indegree and outdegree (neither receive nor send ties). When the indegree and outdegree are equal for all actors, reference is made only to the degree of the actors involved. In the AIDS network everyone has the same indegree as outdegree because of the symmetric nature of sexual ties. In figure 4 or figure 5 one can see that Actor 16 has the highest degree, which is eight. Actors 5 and 26 have equal degrees, because both are linked to exactly five others in the network.

Paths.

A walk of a graph is an alternating sequence of points and lines, not terminating with the beginning node. A trail is a walk whose lines are all different, while a path is a walk whose points are all different. Regarding the AIDS network (Figures 4 and 5), the sequence of ties between actors 37, 36, 31, 32, 31, is a walk, but not a trail, because the tie between actors 31 and 32 is repeated. The sequence of ties between actors 35, 34, 32, 33, 34 is a trail but not a path, because actor 34 is in the series twice. A path is a chain of distinct points, for example the sequence 26, 31, 32, 33, in which no actor is repeated. A path's length is equal to the number of lines which it contains. The distance between any two points $d_i(a, b)$ is the length of the shortest path (geodesic) between the points. A closed path, one which begins

the number of relations received by a given node (matrix column sums), and an outdegree, which is the number of ties a node directs or sends to other nodes (the matrix row sums). Moreno's "stars", then, have a high indegree (receives many ties) and a low outdegree (sends few ties), while the "isolates" have both a zero indegree and outdegree (neither receive nor send ties). When the indegree and outdegree are equal for all actors, reference is made only to the degree of the actors involved. In the AIDS network everyone has the same indegree as outdegree because of the symmetric nature of sexual ties. In figure 4 or figure 5 one can see that Actor 16 has the highest degree, which is eight. Actors 5 and 26 have equal degrees, because both are linked to exactly five others in the network.

Paths.

A walk of a graph is an alternating sequence of points and lines, not terminating with the beginning node. A trail is a walk whose lines are all different, while a path is a walk whose points are all different. Regarding the AIDS network (Figures 4 and 5), the sequence of ties between actors 37, 36, 31, 32, 31, is a walk, but not a trail, because the tie between actors 31 and 32 is repeated. The sequence of ties between actors 35, 34, 32, 33, 34 is a trail but not a path, because actor 34 is in the series twice. A path is a chain of distinct points, for example the sequence 26, 31, 32, 33, in which no actor is repeated. A path's length is equal to the number of lines which it contains. The distance between any two points $d_i(a, b)$ is the length of the shortest path (geodesic) between the points. A closed path, one which begins

and ends with the same node, is a cycle. Deleting actor 35 in the example of the trail above produces the cycle 34, 32, 33, 34. The value of a cycle is calculated by taking the product of the values or signs of its lines. This calculation is used to determine whether or not a cycle is balanced (positive). A negative cycle results from an odd number of negative relations. A positive cycle results if all relations are positive or if there is an even number of negative relations. Knowing the values of all the cycles in graph, one can calculate the net value of a graph at a specific point, P, which is simply the sum of the values of the cycles with which P is a point.

Reachability.

If points a and b are joined by a path, point b is reachable from point a. Reachability (Harary, Norman & Cartwright, 1965) is a measure of the path distance between the nodes of a network. In the AIDS network of Figure 4, actor 5 can *reach* actor 4 and actor 1 with path lengths "1" and "2" respectively. Actor 4 and actor 1 can reach each other in a path length of three. These paths are equal in length due to the symmetric nature of the ties. However, if the ties had been asymmetric, some actors would have been able to reach actors that were incapable of reaching them.

Berkowitz (1982) states that if the reachability of every node in a network is determined, the information can be used to interpret certain characteristics of that network. First, if the range of reachabilities is large, this suggests that there is a difference in the efficiency with which individuals send or receive "information". Second, if the

distribution of reachabilities is skewed, the ability of members to mobilize the resources implicit in the network is also likely to be skewed. Finally, when the reachabilities of a network are low, "misinformation" or "rumor" is likely to be rampant.

Harary, Norman, and Cartwright (1965) proposed a measure of compactness for networks. This is the proportion of nodes that can be reached from any point within a network at a given distance. Compactness is a summary measure of the reachabilities of a network's nodes and reflects the total proportion of nodes which can be possibly reached.

Combining two or more matrices of relations (k_1,k_2,. . . k_n) for the same network of N nodes into one matrix (by matrix addition), indicates mulitplexity, or the existence of multiple relations between nodes. A tie between nodes a and b is said to be *multiplex* if a specific proportion of ties are formed between a and b across k networks. For the AIDS data, three matrices of relations are possible (k = 3). Those relations include location, diagnosis, and sexual ties. After summing the ties across all k networks, the proportion of multiple relations can be determined. At the level of the individual actor, actor multiplexity can be measured as the proportion of an actor's ties with the other N -1 actors across k networks. Actor mulitplexity can be calculated by dividing the actual number of ties by the total number of ties that are theoretically possible, $k(N^2 - N)$. Multiplexity reflects the total proportion of reachable nodes pertaining to multiple relations between a specific set of actors.

Connectedness.

A graph is connected if every pair of points is joined by a path. The connectedness of a pair of points is an indication of how two points may be linked by directed lines. Points which are 3-connected are joined by paths in both directions (symmetric); 2-connected points are joined by a path in one direction only (asymmetric); 1-connected points are joined by lines without regard to the direction of the lines; 0-connected points have no directed lines joining them in either direction and are therefore not reachable (Knoke and Kuklinski, 1982). Harary and Norman (1960) define four degrees of connectedness. A digraph can be strongly, weakly or unilaterally connected or totally disconnected (Figure 6).

Insert Figure 6 here

A digraph is strongly connected if every two points are mutually reachable (3-connected). In the strong digraphs in Figure 6, every point can be reached from every other point because the points complete a cycle. The cycle of actors 34, 32, 33, 34 mentioned previously for the AIDS network is thus strongly connected. Further, if the asymmetric ties of the digraph consisting of points d, e, f, and g were all mutual, including an additional tie between points e and g, it would make the digraph complete. It follows that every complete digraph is strong, but not every strong graph complete.

A digraph is unilaterally connected (unilateral) if, for any two points, at least one is reachable (2 connected). This is easily shown by

the digraph of points d, e, f, g, in Figure 6 -- e is reachable from f, d from e, g from d, but not vice versa. This example illustrates what is meant by strictly unilateral, a digraph that is unilateral but not strong.

A digraph is weakly connected (weak) if there is a sequence between every two points, without regard to direction of the relation (1-connected). For every possible separation of all of the points of a weak digraph into two disjoint, non-empty groups, there must be at least one line having one end point in one group and the other end point in the other group (French, 1956). The digraph consisting of points d, e, f, g, in figure 6 (weak), can be divided so that points d and g are in one group, and points e and g are in another, with the line between points e and d joining these two groups. A weak digraph such as this which is not unilateral, is termed strictly weak.

A graph that is not weak is disconnected (0-connected). A disconnected graph may be separated into two disjoint groups such that no line connects one group to the other. In Figure 6 (disconnected), the digraphs consist of two groups. In the digraph on the left, points a and b form one group, point c, by itself forms another. In the second digraph, points f and g constitute a group as do points e and d. Even though points within each group are connected, there are no ties which connect one group to another, so the digraphs are disconnected.

Density.

The density of a network is the degree of completeness of the ties among the nodes. Density is an indicator of the types of interactions found in the network. As the ties become more dense, the interactions

are likely to be more cliquish, with actors clustering themselves into exclusive groups. A straightforward measure of density is the number of actual ties divided by the total number of ties that theoretically could have been made (N^2 - N, if self-directed loops are not permissible). The density of a network can range between 0.0, a totally disconnected network, to 1.0, a network which is completely connected. For the AIDS network the density is .053 since only 82 of the 1560 possible ties were actually formed.

An actor's "ego-network density" extends the concept of density to the level of the individual actor. The ego-network of an actor consists of only that subset of N - 1 other actors, denoted n_e, with which an actor (ego) has direct relations. This index is calculated in the same manner as network density, but it is the proportion of potential linkages among ego's alters that actually occur (the number of actual linkages divided by n_e^2 - n).

A special type of network density is network cohesion, which takes into account only the mutual ties in a binary matrix of directed ties. It is calculated by dividing the total number of mutual ties by the maximum possible number of such choices (N^2 - N)/2. The network cohesion index ranges from 0.0 to 1.0, with the larger values indicating that a greater proportion of network relations are reciprocated or mutual. Regarding the AIDS network, only 41 mutual ties are actually formed from a possible 780, therefore the network cohesion is .051.

Centrality.

Bavelas (1950) and Leavitt (1951) introduced the idea that the

greater an actor's involvement in all network relations, the more central or accessible that actor is in the network. The simplest measures of centrality are based on an actor's degree. One way of estimating an actor's centrality is to calculate the proportion of relations in which the actor is involved, or more simply the actor's degree divided by the total number of ties in the network. In the AIDS data, the centrality measure for actor 16 would be .097 ($^{8}/_{82}$) because his degree is 8 and there are 82 ties total for the network.

An alternative degree-based method of assessing centrality is to consider an actor's relative degree, which is the ratio of an actor's actual adjacency (degree) to the potential number of first order ties that theoretically could have been made. If the network is of size N, an actor's relative degree is obtained by dividing the actor's degree by N -1. Referring to AIDS' actors 16 and 5 respectively, their relative degrees would be .205 ($^{8}/_{39}$) and .128 ($^{5}/_{39}$), with 8 and 5 referring to their respective degrees and 39 equaling N -1, with 40 actors in the network.

In identifying the potential of certain nodes in a network to be active communicators, degree-based centrality measures are reasonable, because it is assumed that the closer people are in a network, the easier it is to send and receive information. However, degree-based measures of centrality are insufficient for examining the position of nodes within a network. For this reason, structuralists designed two additional types of centrality measures to take the concept of position into account (Freeman, 1979). The position of a node refers to the subgroup in which that node belongs.

The first of these measures is based on the geodesics (shortest paths) linking nodes to one another. Given two nodes a and b, any and all points falling on the shortest paths linking these two nodes are said to fall *between* points a and b. This kind of measure is an index of potential control of communication. Whenever two nodes are not adjacent, but are joined via other points, these points between can distort the flow of communication between these two nodes.

The second type of measures rest on the generalization of the notion of adjacency, the closeness of one point to all others. If any two nodes, a and b, are adjacent, they are said to be joined by a relation at length t^1. Two points, a and c, which are joined via point b are "second-order adjacent" or joined at length t^2; two points joined via two points are t^3, and so on to t^n. The number of indirect linkages between points can be identified by taking powers of the corresponding matrix. Squaring the matrix yields the number of second-order ties, and cubing the matrix results in the number of third-order ties. If the sum of these *orders* of interconnection between any node and all other nodes in a network is calculated, in terms of shortest path distance, an index of independence from control results.

According to Freeman (1979), these kinds of centrality: degree, betweeness, and closeness, "imply three competing theories of how centrality might affect group processes." An effort was needed to sort out the effects of these several types of centrality. In recent years, structuralists have clarified these notions of centrality and have extended them to provide indices of the overall centrality of a network, *centralization*. This index is based on the difference between the

centrality score of the most central actor and that of the N - 1 other actors. Centralization reflects the degree to which a network is dominated by a single point in terms of degree, betweeness, and closeness. The values of centralization range from 0.0, reflecting a network in which no actor is structurally distinct from any other actor (complete graphs), to 1.0, in which a single point dominates the connections among others (network star).

Levels of Analysis

Thus, a network may be analyzed at various levels. The two extremes, the level of the individual actor and the system of actors as a whole, have already been mentioned with respect to several measures. The two remaining, but intermediate, levels of analysis are that of the dyad (pairs of actors) and the triad (sets of three actors). Each of these four distinct levels conveys different information about the network since each differs in the number and complexity of the units under investigation.

The egocentric network is the simplest of these levels. It consists of each individual, all of the actors with which it is adjacent, and the relations between the ego and the adjacent actors. For a network of size N, there are N individual units of analysis. Each actor can be described by the number, magnitude, and other characteristics of linkages with the other actors in a system.

The second level at which a network can be analyzed is the dyad, consisting of a pair of actors. If N nodes compose the network, then there are $(N^2 - N)/2$ dyadic units of analysis. The main concern in

dyadic analysis is whether or not a direct tie exists between the two actors or whether there are some other "higher order", indirect ties through which the actors are connected. Analysis at this level focuses on the function of the joint characteristics of pairs of points.

The triad is the third level of analysis, consisting of sets of three points. For a sample size of N, there are $^1/_6$N(N-1)(N-2) distinct triads formed by selecting each possible subset of three nodes and their linkages. Triad research concentrates mainly on the local structure of sentiment ties among individual actors, with particular concern for determining transitivity relations. A graph is transitive if and only if for nodes x , y, and z in Graph G, when if x chooses y, and y chooses z, then x also chooses z (Holland & Leinhardt,1977). Transitivity can only be studied at the triadic level.

The highest and most important level of analysis is the complete network. It uses the complete information about the patterning of ties among all actors to determine the existence of distinct positions and roles within a system and to describe the nature of the relations among these positons.

Structure of Complete Networks

The objective at the level of the complete network is the determination of a network's structure. The positions or social roles (subgroups within a network) that actors occupy and the patterns of relations among these positions constitute the social structure. The task of the network analyst is to identify these positions. The complexity of the network can be greatly reduced by this procedure

since actors may jointly occupy the same position within a network. Two criteria of major interest, used in the identification of positions and the actors who occupy these positions, are social cohesion and structural equivalence (Burt, 1978).

Social Cohesion and Cliques

By the first criterion, social cohesion, actors are aggregated into positions identified as cliques, in which *all* actors occupying the same postion are directly connected to one another. Many definitions of cliques have been proposed, but most embody the notion that cliques are highly cohesive subsets of actors within a network. Holland and Leinhardt (1977) defined a M-clique (M denotes mutual ties). Within each M-clique pairs are joined by mutual ties, and between two distinct M-cliques all pairs of individuals are either joined by asymmetric or null ties. This is consistent with the most stringent, formal definition of a clique as a maximal complete subgraph, a set of completely linked points not contained within a larger completely linked set (Festinger, 1949; Luce & Perry, 1949). Cartwright and Harary (1956) defined a clique as a maximal subgroup that consists of at least three members. In Rapoport's (1957) single random net, cliques are defined as "independent structures". Davis (1963) describes a clique as a subset of group members whose average liking for each other is greater than the liking for other members of the set.

The completeness criterion for cliques is overly restrictive in the sense that a subset of actors might fail to be identified as a clique because only a few ties among its actors were lacking. An alternative

criterion has been derived from the maximal strong component concept from graph theory. A maximal strong component is a network subgroup in which each actor can reach every other actor directly or indirectly, and no further actors can be added without losing this mutual reachability (Knoke & Kuklinski, 1982). One of the best examples of the application of the maximal strong component definition is on the location of n-cliques (Luce, 1950), which allows for indirect connections but limits the maximum distance at which these interactions can occur.

The k-plex clique (Seidman & Foster, 1978), another relaxation of the stringent clique definition, is a structure with n points in which each point is connected by a path of length "1" to at least n - k of the other points. An actor of a k-plex therefore has maximum strong relations with all except k clique members.

The definition of a k-plex is consistent with the concept of a social circle, in which actors are included in a position if they have a certain percentage (i.e. 85%) of direct linkages with the other position actors. As opposed to a set of distinct cliques, a true social circle should consist of a complex and undivided network of relationships (Crane, 1969). For this reason, the indirect interactions within a network, are an important aspect of social circles. Due to the nature of these structures, their exact boundaries are difficult to locate.

Cartwright and Harary (1956) stated that a s-graph is balanced if all its cycles are positive. In other words, a s-graph is balanced if and only if its points can be separated into two mutually exclusive subsets such that each *positive* line joins two points of the *same* subset and

each *negative* line joins points from *different* subsets. Cartwright and Harary further state that balance implies a tendency for groups to polarize into exactly two cliques.

James A. Davis questioned Cartwright and Harary's partitioning into only two cliques, since research showed that networks could be split into more than just two subgroups. Davis (1967) introduced the idea of clusterability, or multiple clique phenomenon. A clustering of a s-graph, S, is a partition of the point set V(S) into subsets, $P_1, P_2, \ldots, P_n$, or plus sets, such that each positive line joins two points from the same subset and each negative line joins two points from different subsets. S has a unique clustering if and only if S contains no cycle that has exactly one negative line. If this clustering is unique, then its plus-sets are called clusters.

It is important to note the differences between clustering in a complete graph and in an incomplete graph. In a complete clusterable graph, the plus sets are unique, and there is only one way in which they can be formed. In incomplete graphs, however, unique plus-sets are possible, but in most cases there is more than one acceptable way to form plus-sets.

Structural Equivalence

The application of a structural equivalence criterion to a set of relations is the second basic approach to partitioning a network into subgroups. Actors are aggregated into jointly occupied positions if they have a common set of linkages to the other actors in the system. More formally, Lorrain and White (1971) state that two objects a and b

of a set C are *structurally equivalent* if, for any given relation R and any object x of C, aRx if and only if bRx, and xRa only if xRb. In other words, two actors are structurally equivalent if they relate to every object and are related to by every object in a set in exactly the same way. Just as the maximal complete subgraph is an overly restrictive definition for cliques, so is this criterion for structural equivalence overly restrictive. Most researchers relax this criterion, grouping actors into positions on the basis of the similarity of relations between actors.

One of the more obvious contrasts between clique detection methods and the criterion of structural equivalence lies in their differential emphasis on the relations within and between subgroup actors. Structural equivalence procedures do not require any members of the equivalent subset to maintain relations to each other, as cliques do. All that is important are the relations with the other network actors.

A second contrast between methods of clique detection and structural equivalence is that the former is generally applied to single networks, and the latter to multiple networks.

Two of the most popular operationalizations of the structural equivalence approach are measures of continuous and discrete distance. The former method takes the entire network and aggregates individual actors into successfully larger clusters, while the latter method is based on the division of the network into smaller and smaller *blocks*, or sets of structurally equivalent actors. The focus of the rest of this discussion will be on the latter.

Blockmodels.

White, Boorman, and Breiger (1976), introduced the *blockmodel* method of partitioning a network in terms of discrete distance. A blockmodel is a hypothesis about a set of data matrices, specifying for each matrix which blocks will be zeroblocks (blocks with few or no ties) when some common partition is imposed on all matrices (one for each relation) for a specific network of actors. White et al. present five basic ideas which apply to blockmodels. These are: 1) structural equivalence requires that actors of a network be partitioned into distinct sets, each treated homogeneously not only in its internal relations, but also in its relations to each other in such set; 2) the primary indicator of a relation between sets is not the occurence but the absence of ties between individuals in the sets; 3) many different types of tie are needed to portray the social structure of a network; 4) the nature of a type of tie is inferred from the pattern of all ties of that type in a given population; and 5) a social structure model requires specifying whether or not zeroblocks exist. Blockmodeling permits researchers to move beyond the simple configuration of lines and points of graph theory and beyond the inferences that are technically feasible from sociometric diagrams.

White et al. also provide five contrasts between blockmodels and sociometry. First, persons not in cliques are generally disregarded by sociometry, but blockmodels require a complete partition of actors, so that sets of persons can be structurally important regardless of whether or not the sets are cliques. Second, even when cliques are defined by their similarity of ties to third parties rather than by

choices to one another, the clique imagery is retained and often limits interpretation. In blockmodels, partitioning the actors is only one part of the process. The other part is to interpret the pattern formed on the one or more networks by the partition. The third fundamental contrast between sociometry and blockmodels lies in the use of spatial imagery. While sociometry expresses cliques, as well as other concepts, in terms of locations and distances within a space, blockmodels assume no such spatial embedding. The fourth contrast involves boundaries. Sociometry treats the population studied as separated from the rest of the world, whereas blockmodels apply to networks among people who are embedded in a larger world. The final contrast stems from a basic methodological issue of searching for indices of the degree of deviation from classical balance. White et al. argue that sociological analyses need explicit models of structures in observed populations, and blockmodels meet this need.

Blockmodels consist of an abstract pattern among a few aggregate units, a square binary matrix termed an image, that characterizes the more detailed interactions among a larger group of actors. Images are obtained by first rearranging the rows and columns of a matrix (permuting), so that those actors that jointly occupy the same position or block are adjacent in each matrix of the network. These blocks are then reduced to form a binary matrix according to the density of relations in each block. Two methods generally used to accomplish this are 1) coding all blocks with ties among actors as "1" (one-blocks) and those with no ties as "0" (zero-blocks) or 2) apply some cutoff density value, alpha, where all blocks less than the alpha are set to "0" and all

blocks with density greater than or equal to alpha are set to "1".

Computer Algorithms.

Two computer algorithms devised to construct assignments based on structual equivalence are BLOCKER, developed by G. H. Heil (Heil & White, 1974) and CONCOR, developed by Breiger (Breiger, Boorman, & Arabie, 1975). BLOCKER assigns actors to positions so that the rearranged or permuted matrix obeys the given blockmodel. BLOCKER can identify *crystallizers* , actors whose assignment to one or more particular blocks in the blockmodel determines the placement of many of the other actors in the system, and *floaters* , actors which are allowed multiple alternative assignments. CONCOR is a hierarchical clustering algorithm that partitions actors into possible blocks and then finds a blockmodel by inspecting the data matrices rearranged according to the given partition. The difference between these two algorithms is that CONCOR produces from raw data an assignment of individuals to blocks and then suggests a blockmodel hypothesis, while BLOCKER starts with a hypothesis and then derives from it assignment of actors to blocks that satisfies the hypothesis for a given set of data matrices (White et al, 1976). Blockmodels are a natural framework for discussing various changes in a structure, because individual ties can still be consistent with an unchanged structural pattern. Blockmodeling requires that ties of a given type from any actor in one block to any actor in another be equivalent in structural equivalence; however actors do not necessarily have to maintain every tie with all other actors in the same or different positions or blocks.

Models of Analysis

In analyzing the AIDS data, four major models were applied to gain a better understanding of the structure of the network. The knowledge of this structure once acquired can aid in determining the manner in which AIDS is spread. The models include measures based on centrality, structural equivalence, hierarchical clustering, expansiveness, popularity, and reciprocation. Depending on the model, the results may differ, because each is considering a different criterion, so a comparison between models is essential.

Centrality Model

By obtaining measures of centrality for a network including actor centrality measures for each individual, certain information concerning the network is conveyed. This information includes whether or not one actor dominates the network more than any other, or whether all nodes are equal and none dominate. The measures of actor centrality determine which actors are most central and which are more likely to dominate the periphery. From an epidemiological standpoint, those most central to a network are likely to spread a disease because of the important links they provide. By isolating these target actors and vaccinating them first, if the disease has a vaccine, the spread of the disease may be slowed.

Structural Equivalence Model

A second model which may be used to investigate the structure of networks is based on the notion of structural equivalence. To

reiterate, two nodes are structurally equivalent if each node relates to a third node in exactly the same way. As described earlier, BLOCKER and CONCOR are two computer algorithms that are used to assist network analysts in searching for structurally equivalent nodes.

By examining the image of a network, information pertaining to the interactions within and between subgroups of structurally equivalent nodes can be obtained. A subgroup that sends many ties has more "power" or "influence" than a subgroup that neither sends nor receives ties (even to itself), which in this case is a true isolate. On the other hand, a subgroup that has ties almost exclusively with itself is considered a clique.

In terms of studying the spread of disease, investigators should examine the image matrix to pinpoint the subgroups that dominate the matrix in terms of the number of ties sent, and also those subgroups that serve as bridges between other subgroups. Cliques are structures in which the disease is more isolated and less likely to spread to the rest of the network. Identifying true isolates and containing them before any spread is possible would be useful in controlling the disease.

Clustering Model

A third model in the investigation of a network's structure is based on clustering, the multiple clique phenomenon. The clustering of a s-graph was defined earlier as a partition of a point set into plus-sets, such that a positive line joins two points of the same subset, and a negative line joins two points from different sets. Clustering of a

graph or digraph is generally accomplished using distance measures.

UCLUS (D'Andrade, 1978) and COMPLETE LINK (Johnson, 1967) are two algorithms used to examine the clusterability of the AIDS data. UCLUS begins by clustering actors based on a minimum link algorithm. As the program proceeds, pairwise comparisons of clusters (or clusters and actors) are accomplished using the Mann-Whitney U Statistic. The median distances between clusters are ranked, and the clusters with the smallest median distances are merged, until every node in the network is included.

The complete link hierarchical clustering program, also referred to as the maximum distance algorithm, begins by finding pairs of actors that are the least distant and arranging them in the first cluster. The distance between each new cluster and the other actors is defined as the larger of the two distances between each of the two joined clusters and each actor in the matrix. The program searches for compact, isolated clusters of actors, as opposed to a long string of actors found by single link algorithms (minimum distance).

Cluster analysis of a network gives the order in which nodes are likely to group based on their "closeness" to each other in the network. By determining how such groups will cluster in an epidemiological network, cutpoints can be found where a possible vaccination may be implemented. Thereby vaccinating these groups before they can link to other groups might aid in slowing the disease.

p_1 Distribution Model

Traditional methods for the mathematical analysis of digraphs have

been elementary and infrequently utilized multivariate statistical analysis. In the past ten years, the application of log linear models to study complete networks has increased in popularity.

Holland and Leinhardt's (1981) p_1 distribution model is appropriate for the study of single, binary relations defined on a set of nodes; that is univariate digraphs with no data on nodal attributes. This model, as well as others of its kind, places a probability function on the ties between actors in a network by specifying the probability that a pair of actors will form one of the four possible dyadic relations (mutual, "sender" asymmetric, "receiver" asymmetric, or null). Assuming dyads are independent the probability distribution can be obtained by multiplying the dyad's probabilities to procure their joint distribution.

Fienberg and Wasserman (1981) describe the four sets parameters employed in the p_1 model: $\{\alpha_i\}$, $\{\beta_j\}$, $\{\rho\}$, and $\{\lambda_{ij}\}$. The $\{\alpha_i\}$ parameters measure the production of actors, reflecting the probability that an actor will *send* ties. The $\{\beta_j\}$ parameters measure the popularity of actors, or how likely an actor is to *receive* ties. The likelihood of reciprocated ties is measured by $\{\rho\}$ parameters. The $\{\lambda_{ij}\}$ parameters refer to "dyadic effects" that are present in the data structure to insure that certain constraints hold. Fitting the p_1 model to an "observed" data matrix is equivalent to constructing an "expected" sociomatrix with indegrees, outdegrees, number of mutual ties, and the total number of the choices identical to those of the observed sociomatrix. If the difference between the observed and expected matrices is large, then a more complex model is needed which can take

into account additional structural effects.

In order to fit the p_1 model to a set of data, a y-array is constructed. A y-array, an expanded version of the original matrix, may be either three-dimensional (undirected graph) or four-dimensional (directed graph). For a symmetric sociomatrix with N nodes and a relation with C strengths (C = 2 for binary data), the y-array has the dimensions N x N x C. An asymmetric sociomatrix has a y-array with dimensions N x N x C x C. The reason for this difference in dimensions is that a tie from i to j, denoted X_{ij}, is not the same as a tie from j to i, X_{ji}, in an asymmetric matrix, so an additional dimension is necessary to describe the data.

The y-array corresponding to the asymmetric digraph in Figure 1b. is shown in the top half of Table 2. For the symmetric ties in the graph, where $X_{ij} = X_{ji} = 1$ a "1" is placed in the second row, second column of

---

Insert Table 2 here

---

the smaller 2 x 2 matrix, or in the first row, first column if $X_{ij} = X_{ji} = 0$. For an asymmetric tie such that $X_{ij} = 1$ and $X_{ji} = 0$, a "1" is placed in the second row, first column of the 2 x 2 matrix, or in the first row, second column if $X_{ij} = 0$ and $X_{ji} = 1$. Notice that by constructing the y-array in this manner, the matrix remains binary. However, in constructing this matrix the size increases rapidly as N becomes larger. It is for this reason that the y-array for the AIDS data cannot be shown since its dimensions are 40 x 40 x 2, or 3200 cells.

To study nodal attribute variables, the y-array is aggregated across cells to form a w-array. The original data are grouped by an attribute that characterizes the data (such as sex, race, or religion), and the y-array permuted so that the nodes in each subgroup, denoted S(i), are adjacent. The nodes in each subgroup are considered as a "block", and the y-array is aggregated according to these blocks. A w-array corresponding to an asymmetric y-array has dimensions S x S x C x C, where S equals the number of subgroups formed. The w-array created from the y-arrray example mentioned previously is shown in the bottom of Table 2. Actors a, b, and c are grouped as S(1), and actors d and e are grouped as S(2). This division was arbitrary and for the reader's convenience to see how the two arrays correspond. The dimensions for this w-array are 2 x 2 x 2 x 2. Notice that the w-array no longer represents binary data.

A symmetric y-array will produce a w-array with dimensions S x S x C. For the AIDS data, two nodal attribute variables were available, location and diagnosis. Both sets of variables could be partitioned into three subgroups. These w-arrays are shown in Table 3.

Insert Table 3 here

The location variable was divided by residents residing in California, S(1), New York, S(2), and Elsewhere, S(3), (Pennsylvania, Georgia, New Jersey, Texas, and Florida). Similarly, the diagnosis variable was divided by those who had Kaposi's Sarcoma, S(1), Pneumocystis Carnii

Pneumonia, S(2), or another infection, S(3), (Multiple diagnosis or other opportunistic infection). The dimensions for both the w-arrays are 3 x 3 x 2. Note the difference between these w-arrays and the w-array for the previous example. The cells in these matrices are the sums of the zeros and the ones for each "block" of actors in the y-array. For example, the second cell in each block, denoted $X_{ji} = 1$, is the total number of ties that were formed in that block, and the first cell, $X_{ji} = 0$, is the number of ties that were null. The blocks on the diagonal represent the total number ties for each subgroup, while the off-diagonal blocks represent the total number of ties between subgroups.

By fitting a log linear model to a y-array or w-array, parameter estimates for each individual or subgroup can be obtained. The number of parameter estimates depends on the model used. An asymmetric y-array can be represented by a full model, one with all parameters included. A symmetric y-array, can be represented by either the $\{a_i\}$ or the $\{b_j\}$ parameter since, in this case, these parameters are equivalent. These equal parameters are called $\{\gamma_i\}$ parameters. This is a very simplified explanation of the models and is not intended to be complete, but only to show the difference between the asymmetric and symmetric data models. For a more detailed description of the types and various forms of log linear models, see Fienberg and Wasserman (1981).

By examining the parameter estimates for each individual or subgroup, one can determine the independent effects of expansiveness,

popularity, and reciprocating behavior for each node or group. These parameter estimates give more information about the nodes in a network than do measures of degree. The parameter estimates determine which nodes will be more expansive, popular, or more reciprocating. The higher the parameter estimate the more likely that type of behavior. For instance, a node with $\{\alpha_i\}$ equal to .430 is more likely to form ties than a node with an $\{\alpha_j\}$ equalling .019. The same is true for subgroups, except the parameter estimates do not refer to individuals within the group, but the group itself.

By obtaining parameter estimates for every actor, the actors can be divided into stochastically equivalent groups. By stochastically equivalent it is meant that for every node in a group, the $\{\alpha_i\}$'s are constant, as well as the $\{\beta_j\}$'s. In the case of the AIDS data all nodes with the same $\{\gamma_j\}$'s would be in the same group.

Method

Subjects

The subjects for this study consisted of forty patients with the Acquired Immune Deficiency Syndrome linked by sexual contact. The data and information concerning these patients were provided by Auerbach et al. (1984). These data included information concerning the sequence of onset, diagnosis, and the state of residence of each patient at the time of onset.

Materials

The data were analyzed using UCINET (MacEvoy & Freeman, 1987), a microcomputer package for network analysis. UCINET includes over forty network programs integrated into eight menu-driven modules. These modules include: DATA, for creating data matrices; GRAPH, for computing graph-based distance and centrality measures; PA (positional analysis), for computing measures of similarity among actors; CLUS (cluster), for computing cluster analyses from similarity or dissimilarity measures; and SUBG (subgroups), for finding cliques and clique-like structures in a network.

Procedure

The AIDS network provided graphically by Auerbach et al. (1984) was translated into a square (40 X 40), binary, symmetric matrix. This (40 X 40) matrix was examined to obtain various information and measures, including regular and structural equivalence,

hierarchical clustering, reachability, and number of geodesics for each actor.

To study the onset of the disease AIDS, the original (40 X 40) matrix was permuted to construct several smaller matrices. The first matrix consisted of only those patients in the network who were first diagnosed as having AIDS. The second, consisted of those patients in the first matrix in addition to those patients who were sequentially diagnosed next, and so on. Each of these matrices were analyzed with respect to actor centrality measures, structurally equivalent groups, hierarchical clustering, and number of geodesics for each actor.

The original data set was transformed into a (40 X 40 X 2) y-array and a log linear model applied using SPSSX to obtain parameter estimates, $\{\hat{\gamma}_i\}$'s, for each actor. The actors were divided according to their parameter estimates into stochastically equivalent groups. The original data set was then grouped by two attribute variables, location and diagnosis. The y-array was aggregated across cells according to these subgroups and two (3 x 3 x 2) w-arrays were created. A log linear model was applied using SPSSX for both w-arrays and parameter estimates, $\{\hat{\gamma}_i\}$'s, obtained for every subgroup.

Results and Discussion

Centrality Model

Due to the relatively large size of the data matrix and the limits of the program, UCINET could only give actor centrality measures for 32 of the 40 nodes. The decision on which nodes to include and which to cut was finally based on the sequence of disease onset. The last eight nodes to be diagnosed, 21, 23, 24, 26, 27, 34, 37, and 40, were not included in the analysis, and actor centrality measures were found for the remaining 32 actors. See Table 4 for the rankings of the the actors for both betweeness and relative degree centrality measures. The results of the measures differ because of the different criterians used for each.

Insert Table 4 here

Node 16 ranked first on both measures of centrality. For the remaining 31 actors, four tiers of actors can be distinguished. Each tier represents actors who as a group ranked the same on both betweeness and relative degree. The first tier includes nodes 5, 8,11, and 20. As a group these nodes ranked second, but their individual ranks varied slightly within the group. These nodes either had direct links with actor 16 or were a path length of 2 away from the central node. The second tier nodes, 2, 9, 14, 19, 22, and 28, are ranked third as a group, and their individual ranks also varied

within the group across both measures. Some of these nodes also had direct links to actor 16, but were not included in tier one because of their different linking structure. All those in tier one are linked to at least three other actors, whereas all those in tier two are linked to only two other actors. The third tier includes nodes 29, 31, 32, 1, 3, 4, 6, 7, 10, 12, 13, 15, 17, 18, 25, 30, 33, 33, and 36. These actors are individually ranked the same on both measures, and are the actors that characterize the periphery. The fourth tier, 38, 39, and 35, represent the isolates. This tier is ranked last, and the individual ranks vary only slightly within the group.

These results would have differed if all forty nodes had been included. There would have been no isolates, since node 26 connects the "isolates" to the rest of the group making the network a complete graph. Because of the position of node 26 in the network it would have a relatively high ranking. Other nodes such as 22, 28, and 31, and the nodes connected to these would be similarly affected.

Structural Equivalence Model

CONCOR was used to investigate the structural equivalence of the actors. The network was partitioned into four blocks. After the first split, the nodes were divided into two groups, g_1 and g_2, N = 22 and N = 18, respectively. On the second split, 8 nodes were separated from g_1 to form g_3, and g_2 remained the same. The third split divided the nodes of g_2 into two groups of size N = 7 and N = 11, and g_1 and g_3 remained the same.

The final block partitions are shown in the top of Table 5.

Insert Table 5 here

The first block consists of actors 1, 5, 8, 10, 13, 16, 18, 19, 23, 25, 26, 29, 39, and 40. The second block included only the nodes 2, 3, 4, 6, 7, 9, and 11. The nodes in the third and fourth blocks are 30, 31, 32, 33, 34, 35, 36, 37, and 12, 14, 15, 17, 20, 21, 22, 24, 27, 28, and 38, respectively.

The densities of the ties of each block with itself and the other blocks were calculated and are shown in the middle of Table 5. All those blocks with densities ≤ .09 were set to zero, and all those blocks with densities >.09 were set to one. The resulting image matrix is shown in the bottom of Table 5. Examination of the image shows that the nodes in block 3 only relate to other nodes in block 3, indicating that this block represents a clique-like structure. The other three blocks, 1, 2, and 4, have the majority of their ties with other blocks, but not with themselves. It is important to note that blocks 2 and 4 only have relations with block 1. This block of actors serves as a bridge for blocks 2 and 4. Without this block, actors in blocks 2 and 4 would not be reachable. Structures such as block 1 would be a starting point for researchers to test possible vaccinations to prevent further spread of the disease.

Clustering Model

The results for both clustering analyses, using UCLUS and the COMPLETE LINK algorithm, were fairly similar. Both algorithms divided the nodes into 7 initial groups. Figure 7 shows the differences in node assignments to these initial groups and the differences in which each algorithm hierarchically clustered the network. UCLUS

Insert Figure 7 here

---

clustered the middle and bottom regions of the network before clustering these with the top. The COMPLETE LINK algorithm clustered the top and middle regions separate form the bottom region before all portions of the network clustered together.

Nodes 1, 2, 3, 4, 5, 6 as well as nodes 22, 23, 24, and 25 are grouped the same by both algorithm. The remaining clusters only differ with respect to a few actors, 11, 18, 20, 29, 30, 31, 36, and 37. These nodes represent points in the network which could belong to either cluster. Examining the behavior of these nodes, it is found that these nodes tend to "jump" from one cluster to another in subclusters. This is true for (31, 36, 37), (18, 20), and (29, 30), except for node 11 which is a definite cutpoint in the network. These "jumping" nodes are peripheral points to the clusters themselves and form links between clusters, and if detected, are useful cutpoints separating one cluster from another.

p_1 Distribution Model

The parameter estimates, $\{\hat{\gamma}_i\}$'s, for each individual, obtained using SPSSX, are given in Table 6. The actors are divided into 6 subgroups according to their stochastical equivalence.

Insert Table 6 here

Examination of the data shows that the degree of the node is postively related to the value of the given parameter estimate. The largest subgroup, N = 20, has the lowest $\{\hat{\gamma}_i\}$'s (-.590). This subgroup characterizes the periphery with each node having degree equal to 1. These actors are the least likely to form ties. The second and third sets, each of size N = 8, have degree 2 and 3 respectively. The $\{\hat{\gamma}_i\}$'s equal .128 for the second subgroup and .682 for the third subgroup. The remaining four nodes constitute the next three subgroups. These subgroups have the highest parameter estimates and are more likely to form ties than the first three sets of nodes already described. Nodes 5 and 26, each of degree 5, have $\{\hat{\gamma}_i\}$'s equal to 1.256. Nodes 22 ($\hat{\gamma}_i$ = .936) and 16 ($\hat{\gamma}_i$ = 1.830) are unique to the network because both these nodes occupy a position entirely to themselves.

The parameter estimates for the subgroups in each w array (location and diagnosis) are shown in Table 7. These parameter estimates give an indication of how the nodes behave as a subgroup.

Insert Table 7 here

The actors from New York ($\hat{\gamma}_{ny}$ = .166), as a subgroup, are more likely to form ties than individuals from California ($\hat{\gamma}_{c}$ = -.244) or anywhere else ($\hat{\gamma}_{e}$ = .078).

By diagnosis individuals stricken with Pneumocystis Carnii Pneumonia ($\hat{\gamma}_{p}$ = .152) are more likely to form ties, as a subgroup, than those with Kaposi's Sarcoma ($\hat{\gamma}_{ks}$ = -.156) or other infection ($\hat{\gamma}_{oi}$ = .004). It can be shown from this example, that the subgroup with the most nodes, Kaposi's Sarcoma, does not necessarily have the highest parameter estimate. In fact, this subgroup has the lowest $\{\hat{\gamma}_i\}$'s of all three subgroups, while Pneumocystis Carnii Pneumonia has the highest $\{\hat{\gamma}_i\}$'s, and contains the fewest number of nodes.

Points in the network in which the parameters are the highest are the areas in which more ties are likely to form. These areas which have the highest probability of spreading are points for a possible vaccination to be introduced. For this reason, nodes such as 16, 5, 26, and 22, may be of special interest to researchers who investigate the spread of AIDS.

Comparison of Models

Although centrality measures for all 40 nodes could not be obtained using UCINET, the values of betweeness and relative degree for each

actor were calculated and given in Stephenson & Zelan (1989). Their rankings of each node by relative degree and the rankings obtained using the p_1 distribution model were exactly the same. This results from the fact that the parameter estimates, $\{\hat{\gamma}_i\}$'s, given for each actor were postively related to the degree of that actor, and therefore the relative degree. Recall that the relative degree is the number of actual ties formed divided by N - 1.

Comparisons between CONCOR and the COMPLETE LINK algorithm showed that some of the nodes tended to "cluster" together. These "clusters" included: (1, 5), (8, 10), (13, 16), (19, 26, 40), (23, 25), (22, 24), (27, 38), (2, 3, 4, 6), (7, 9, 11), (31, 36, 37), (32, 33, 34, 35), and (12, 14, 15, 17, 20, 21). A similar comparison was done between CONCOR and UCLUS. The "clusters" found in this comparison were: (1, 5), (8, 10), (13, 16), (18, 19), (23, 25), (26, 29, 39, 40), (2, 3, 4, 6), (7, 9), (31, 32, 33, 34, 35, 36, 37), (12, 14, 15, 17, 21), (22, 24), and (27, 38). The intersection of these two sets shows the nodes which are structurally equivalent within sets of nodes that generally cluster together. These are: (1, 5), (8, 10), (13, 16), (23, 25), (2, 3, 4, 6), (27, 38), (7, 9), (31, 36, 37), (32, 33, 34, 35), (12, 14, 15, 17, 21), and (26, 39, 40). A graphic representation of these results are shown in Figure 8.

Insert Figure 8 here

A comparison between UCLUS, COMPLETE LINK algorithm, and the p_1

distribution model, accomplished in a similar manner as above, shows the nodes that are stochastically equivalent within sets of nodes that generally cluster. These nodes are: (1, 3, 4, 6), (7, 10), (39, 27, 40), (12, 13, 15, 17, 20), (24, 25), and (32, 34). A further comparison between this set of nodes and the set found above (intersection between UCLUS, COMPLETE LINK, and CONCOR) show that the sets (3, 4, 6), (32, 34), (39, 40), and (12, 15, 17, 21) are stochastically and structurally equivalent within the same cluster. These results are also shown in Figure 8.

The nodes found to be stochastically and structurally equivalent within the network were found by comparing the results of the p_1 distribution model with the results from CONCOR. These were found to be: (5, 26), (19, 23, 29), (1, 10, 13, 18, 25, 39, 40), (2, 9), (3, 4, 6, 7), (31, 32, 34), (33, 36), (12, 15, 17, 21, 24, 27), and (20, 28, 38).

Conclusion

The analysis of the AIDS network by Auerbach et al. (1984) found support for the infectious agent hypothesis and demonstrated that sexual relations allowed the transmission of the disease. This analysis was completed prior to the discovery that the Human Immunodeficiency Virus (HIV) was spread by the exchange of bodily fluids.

Network Analysis can also be used to predict the rate and extent of the disease. The detection of certain structures within the network as well as the application of several measures derived from graph theory, can aid in this prediction. The more cliquish the ties, the more likely the disease is to be contained within those cliques, and not to the rest of the network. The density and connectedness of ties within the network are positively related to the spread, while the reachability is postively related to the extent of the disease. The higher the density and connectedness measures, the more likely the spread, while the higher the reachabilty the greater the extent of the disease. Although every patient in the AIDS data was reachable, the network itself was not highly connected, and the density was only .053. This suggests that although the disease was extensive, the progression through the network was probably slow.

An image matrix for a network can show how blocks of individuals relate to each other. Examining the image of the AIDS data, it was found that the individuals in Block 1 served as a bridge for the ties between Blocks 2 and 4. Without Block 1, the individuals in Blocks

2 and 4, would have been unreachable. Detection of structures such as Block 1 and further intervention may aid in slowing the rate and extent of the spread, while detection of cliques (Block 3) may help in isolating the disease.

Cluster analyses of the network were fairly consisent, except for a few nodes (11, 18, 20, 29, 30, 31, 36, 37) that were located on the periphery of the clusters and tended to cluster together. Further analysis of the way these nodes cluster shows that every node within each cluster had the same diagnosis and resided in the same approximate location. These represent people who form important links between clusters of individuals. An example of such individuals are female and male prostitutes. If identified and the proper action taken (given a possible vaccination or merely education), the disease can be stopped from entering certain subpopulations that have not been infected yet.

This study also found that a person's centrality (based on degree) was directly related to the probability of that person spreading the disease. Therefore individuals most central to the network (i.e. nodes 16, 5, 26, 22) are more likely to spread the disease than individuals less central to the network, or those located on the periphery. Analyses by diagnosis and location showed that those individuals from New York or who were diagnosed as having Pneumocystis Carnii Pneumonia may also have a higher probability of spreading the disease, but this result may be specific to the data.

The forty AIDS victims described in this paper have probably already died, as well as thousands of more. It is important that this disease is

studied and ways to halt the spread of the disease found. AIDS has had a profound effect on society as well as those individuals affected. The society as a whole is starting to change their behavior: safe sex is being taught, condoms are advertised on television, intravenuous drug users are warned of the dangers of sharing needles, and blood screening for the Human Immunodeficiency Virus (HIV) by hospitals and clinics is now routine. A cure needs to be found, but until then, understanding the spread of the disease by studying AIDS networks can give researchers some of the answers they need.

Bibliography

Auerbach, D. M., Darrow, W. W., Jaffe, H. W., & Curran, J. W. (1984). Cluster of cases of the acquired immune deficiency syndrome. The American Journal of Medicine, 76, 487-492.

Bavelas, A. (1950). Communication patterns in task oriented groups. Journal of Acoustical Society of America, 22, 271-282.

Berkowitz, S. D. (1982). An introduction to structural analysis: network approach to social research. Toronto, Canada: Butterworth and Company.

Bott, E. (1955). Urban families: conjugal roles and social networks. Human Relations, 8, 345-383.

Breiger, R. L., Boorman, S. A., & Arabie P. (1975). An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling. Journal of Mathematical Psychology, 12, 328-383.

Burt, R. S. (1978). Cohesion versus structural equivalence as a basis for network subgroups. Sociological Methods and Research, 7, 189-212.

Burt, R. S., Christman, R. P., & Kilburn, H. C. (1980). Testing a structural theory of corporate cooptation: interorganizational directorate ties as a strategy for avoiding market constraints on profits. American Sociological Review, 45, 821-841.

Cartwright, D. & Harary, F. (1956). Structural balance: a generalization of Heider's theory. Psychological Review, 63, 277-292.

Cook, K. S., & Emerson, R. M. (1978). Power equity and committment

in exchange networks. American Sociological Review, 25. 722-727.

Crane, D. (1969). Social structure in a group of scientists: a test of the "invisible college" hypothesis. American Sociological Review. 34. 335-352.

D'Andrade, R. (1978) U-statistic hierarchical clustering. Psychometrika, 43, 59-67.

Davis, J. A. (1963) Structural balance, mechanical solidarity, and interpersonal relations. American Journal of Sociology. 68, 444-463.

Davis, J. A. (1967). Clustering and balance theory in graphs. Human Relations, 20, 181-187.

Festinger, L. (1949). The analysis of sociograms using matrix algebra. Human Relations. 2. 153-158.

Fienberg, S. E. & Wasserman, S. S. (1979) Categorical data analysis of single sociometric relations. In S. Leinhardt (Ed.) Sociological Methodology 1981. (pp. 156-192). San Francisco: Jossy-Bass.

Freeman, L. C. (1979). Centrality in social networks. I. conceptual clarification. Social Networks. 1, 215-239.

French, John R. P. (1956). A formal theory of social power. Psychological Review. 63. 181-194.

Granovetter, M. (1974). Getting a job: a study of contacts and careers. Cambridge, Mass: Harvard University Press.

Hallinan, M. T. (1974) The structure of positive sentiment. New York: Elsevier.

Harary, F. (1959) Graph theoretic methods in the management

in exchange networks. American Sociological Review, 25, 722-727.

Crane, D. (1969). Social structure in a group of scientists: a test of the "invisible college" hypothesis. American Sociological Review, 34, 335-352.

D'Andrade, R. (1978) U-statistic hierarchical clustering. Psychometrika, 43, 59-67.

Davis, J. A. (1963) Structural balance, mechanical solidarity, and interpersonal relations. American Journal of Sociology, 68, 444-463.

Davis, J. A. (1967). Clustering and balance theory in graphs. Human Relations, 20, 181-187.

Festinger, L. (1949). The analysis of sociograms using matrix algebra. Human Relations, 2, 153-158.

Fienberg, S. E. & Wasserman, S. S. (1979) Categorical data analysis of single sociometric relations. In S. Leinhardt (Ed.) Sociological Methodology, 1981. (pp.156-192). San Francisco: Jossy-Bass.

Freeman, L. C. (1979). Centrality in social networks. I. conceptual clarification. Social Networks, 1, 215-239.

French, John R. P. (1956). A formal theory of social power. Psychological Review, 63, 181-194.

Granovetter, M. (1974). Getting a job: a study of contacts and careers. Cambridge, Mass: Harvard University Press.

Hallinan, M. T. (1974) The structure of positive sentiment. New York: Elsevier.

Harary, F. (1959) Graph theoretic methods in the management

sciences. Management Sciences, 5, 387-403.

Harary, F. & Norman, R. Z. (1953). Graph theory as a mathematical model in the social sciences. Ann Arbor: Institute for Social Research.

Harary, F., Norman, R. Z. & Cartwright, D. (1965). Structural Models: an introduction to the theory of directed graphs. New York: Wiley.

Heider, F. (1946) Attitudes and cognitive organization. Journal of Psychology, 21, 107-112.

Heil, G. H., and White, H. C. (1974). An algorithm for constructing homomorphisms of multiple graphs. Unpublished paper, Department of Sociology, Harvard University.

Holland, P. W., & Leinhardt, S. (1971). Transitivity in structural models of small groups. Comparative Group Studies, 2, 107-124.

Holland, P. W., & Leinhardt, S. (1981). An exponential family of probability for directed graphs. Journal of American Statistical Association.

Johnson, S. (1967). Hierarchical clustering schemes. Psychometrika, 32, 241-253.

Klinger, R. J. & Minton, J. P. (1973). Case clustering of Hodgkin's disease in a small rural community, with associations among cases. Lancet, 1, 168-171.

Klovdahl, A. S. (1985). Social networks and the spread of infectious diseases: the AIDS example. Social Science & Medicine, 21, 1203-1216.

Klovdahl, A. S. (1986). View_ Net: a tool for network analysis. Social Networks, 8, 313-342.

Knoke, D. & Kuklinski, J. H. (1982). Network Analysis. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-028, Beverly Hills and London: Sage Publications.

Leavitt, H. J. (1951). Some effects of communication problems on group performance. Journal of Abnormal Psychology, 46, 38-50.

Leinhardt, S. (1977). Social networks: a developing paradigm. New York: Academic Press.

Lorrain, F., & White, H. C. (1971). Structural equivalence of individuals in social networks. Journal of Mathematical Sociology, 1, 49-80.

Luce, R. D. (1950). Connectivity and generalized cliques in sociometric group structure. Psychometrika, 15, 169-190.

Luce, R. D. & Perry, A. (1949). A method of matrix analysis of group structure. Psychometrika, 14, 94-116.

MacEvoy, B. & Freeman L. (1987) UCINET: a microcomputer package for network analysis. Mathematical Social Science Group, School of Social Sciences. University of California at Irvine.

Mariolis, P. (1975). Interlocking directorates and control of corporations. Social Science Quarterly, 56, 425-439.

Mitchell, J. C. (1969) The concept and use of social networks. In J. Clyde Mitchell (Ed.) Social Networks in Urban Situations. (pp. 1 - 50) Manchester, England: University Press.

Moreno, J. L. (1934) Who shall survive?. Washington, D. C.: Nervous and Mental Disease Monograph 58.

Paine, L. (1988). AIDS, Psychiatric and Psychosocial Perspectives. Croom Itelv, ltd, Prudent House.

Pearson, H. E. & Rendtroff, R. C. (1945). Studies of the distribution of poliomyelitus virus. American Journal of Hygeine, 41, 179-187.

Rapoport, A. (1957). Contribution to the theory of random and biased nets. Bulletin of Mathematical Biophysics, 19, 257-277.

Rogers, E. M. & Kincaid, D. C. (1981). Communication networks: toward a new paradigm for research. New York: Macmillian.

Schimpff, S. C., Schimpff, C. R., Brager, D. M., & Weirnik, P. H. (1975). Leukemia and lymphoma patients interlinked by prior social contact. Lancet, 1, 124-129.

Seidman, S. B. & Foster, B. L. (1978). A graph-theoretic generalization of the clique concept. Journal of Mathematical Sociology, 6, 139-154.

Selik, R., Haverkos, H. W., & Curran, J. W. (1984). Acquired immune deficiency syndrome (AIDS) trends in the United States, 1971-1982. American Journal of Medicine, 76, 493-500.

Stevenson, K., & Zelan, M. (1989). Rethinking Centrality: Methods and Applications. Social Networks to appear.

Tyckoson, D. A. (1988) AIDS 1987, Ornyx Science Bibliographies: Orynx Press.

Vianna, N. J., Greenwald, P., & Davies, J. P. N. (1971). Extended epidemic of Hodgkin's disease in high-school students. Lancet, 1, 1209-1211.

Vianna, N. J., Greenwald, P., Brady, J., Polan, A. K., Dwork, A., Mauro, J. & Davies, J. N. P. (1972). Hodgkin's disease: cases with features of a community outbreak. Annals of Internal Medicine, 77, 169-180.

Wasserman, S. S. (1978). Models for binary directed graphs and their

applications. Advances in Applied Psychology, 10, 803-818.

Wellman, B. (1983). Network analysis: some basic principles. Sociological Theories, 1, 155-200.

White, H. C., Boorman, S. A. & Brieger, R. L. (1976). Social structure from multiple networks. I. Blockmodels of role and positions. American Journal of Sociology, 81, 730-780.

Table 1

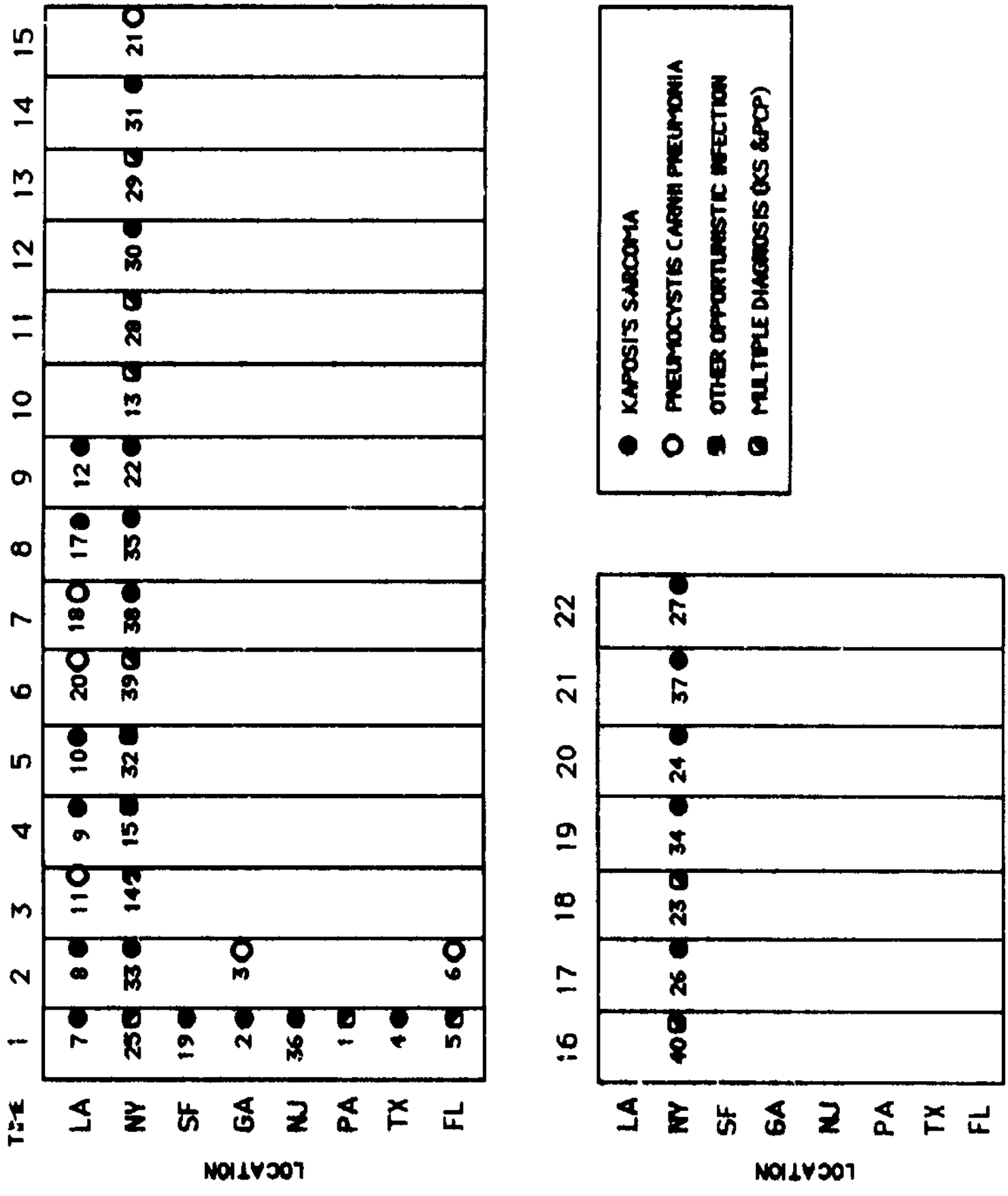

KAPOSI'S SARCOMA
PNEUMOCYSTIS CARINII PNEUMONIA
OTHER OPPORTUNISTIC INFECTION
MULTIPLE DIAGNOSIS (KS &PCP)
LOCATION
LA
NY
SF
GA
NJ
PA
TX
FL
TIME
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
16 17 18 19 20 21 22

Table 2

			j:	a		b		c		d		e	
			Xji =	0	1	0	1	0	1	0	1	0	1
i:	a	Xij = 0		0	0	0	0	0	1	1	0	0	0
		Xij = 1		0	0	0	1	0	0	0	0	1	0
	b	Xij = 0		0	0	0	0	0	0	1	0	1	0
		Xij = 1		0	1	0	0	1	0	0	0	0	0
	c	Xij = 0		0	0	0	1	0	0	0	1	0	0
		Xij = 1		1	0	0	0	0	0	0	0	0	1
	d	Xij = 0		1	0	1	0	0	0	0	0	0	1
		Xij = 1		0	0	0	0	1	0	0	0	0	0
	e	Xij = 0		0	1	1	0	0	0	0	0	0	0
		Xij = 1		0	0	0	0	0	1	1	0	0	0

		S(j) =	1		2	
		Xji =	0	1	0	1
S(i) =	1	Xij = 0	0	2	3	1
		Xij = 1	2	2	1	1
	2	Xij = 0	3	1	0	1
		Xij = 1	1	1	1	0

S (1) = { a b c} S(2) = {d e}

Table 3

Location		S(j) = 1		S(j) = 2		S(j) = 3	
		Xji = 0	Xji = 1	Xji = 0	Xji = 1	Xji = 0	Xji = 1
S(i) =	1	46	10	75	5	170	6
	2	75	5	78	12	219	1
	3	170	6	219	1	426	36

S(1) = California

S(2) = New York

S(3) = Elsewhere

Diagnosis		S(j) = 1		S(j) = 2		S(j) = 3	
		Xji = 0	Xji = 1	Xji = 0	Xji = 1	Xji = 0	Xji = 1
S(i) =	1	476	30	133	5	237	16
	2	133	5	26	2	63	3
	3	237	16	63	3	106	2

S(1) = Kaposi's Sarcoma

S(2) = Pneumocystis Carnii Pneumonia

S(3) = Other

Table 4

	Centrality Measures	
Rank	Betweeness	Degree
1	16 (204.0)	16 (.226)
2	11 (164.0)	5 (.161)
3	5 (104.0)	8 (.097)
4	20 (99.0)	11 (.097)
5	8 (65.0)	20 (.097)
6	19 (63.0)	2 (.065)
7	28 (44.0)	9 (.065)
8	2 (23.0)	14 (.065)
9	9 (23.0)	19 (.065)
10	14 (23.0)	22 (.065)
11	22 (23.0)	28 (.065)
12	29 (23.0)	29 (.065)
13	31 (2.0)	31 (.065)
14	32 (2.0)	32 (.065)
15	1 (0.0)	1 (.032)
16	3 (0.0)	3 (.032)
17	4 (0.0)	4 (.032)
18	6 (0.0)	6 (.032)
19	7 (0.0)	7 (.032)
20	10 (0.0)	10 (.032)
21	12 (0.0)	12 (.032)
22	13 (0.0)	13 (.032)
23	15 (0.0)	15 (.032)
24	17 (0.0)	17 (.032)
25	18 (0.0)	18 (.032)
26	25 (0.0)	25 (.032)
27	30 (0.0)	30 (.032)
28	33 (0.0)	33 (.032)
29	36 (0.0)	36 (.032)
30	38 (0.0)	38 (.032)
31	39 (0.0)	39 (.032)
32	35 (0.0)	35 (.000)

Table 5

Block	N	Nodes
1	14	1 5 8 10 13 16 18 19 23 25 26 29 39 40
2	7	2 3 4 6 7 9 11
3	8	30 31 32 33 34 35 36 37
4	11	12 14 15 17 20 21 22 24 27 28 38

Block Densities

BLOCK =1	.000	.112	.018	.136
2	.112	.000	.000	.000
3	.018	.000	.250	.000
4	.136	.000	.000	.000

Image Matrix

BLOCK =1	0	1	0	1
2	1	0	0	0
3	0	0	1	0
4	1	0	0	0

Table 6

$\hat{\gamma}_i$	n	Nodes
-.590	20	1 3 4 6 7 10 12 13 15 17 18 21 24 25 27 30 35 37 39 40
.128	8	2 9 14 19 23 29 33 36
.682	8	8 11 20 28 31 32 34 38
.936	2	22
1.256	1	5 26
1.830	1	16

Table 7

	$\hat{\gamma}_{s(l)}$
Location	
California	- .244
New York	+.166
Elsewhere	+.078
Diagnosis	
Kaposi's Sarcoma	-.156
Pneumocystis Carnii Pneumonia	+.152
Other	+.004

Figure Captions

Figure 1. Comparison of a graph (1a.) and digraph (1b.).
Reprinted from Harary, F. (1959). Graph theoretic methods in the management sciences. Management Sciences, 5, 387-403.

Figure 2. Comparison of a s-graph (2a.) and a s-digraph (2b.).
Derived from graphs in Figure 1.

Figure 3. Matrices corresponding to the relations shown in Figure 1a. and 1b. respectively.
Matrix 3a. derived from figure 1a.; Matrix 3b. reprinted from Harary (1959).

Figure 4. Graphic represention of the sexual links between 40 AIDS patients including diagnosis and location. The nodes are numbered 1 - 40 arbitrarily and do not correspond to disease onset.
Auerbach, D. M., Darrow, W. W., Jaffe, H. W., & Curran, J. W. (1984). Cluster of cases of the acquired immune deficiency syndrome. The American Journal of Medicine. 76. 487-492.

Figure 5. Matrix representation of the AIDS data. Derived from Figure 4. The numbers correspond to the values given the nodes in Figure 4.

Figure 6. Digraphs separated into exclusive connectedness categories- strong, unilateral, weak, and disconnected.

Figures on the left--reprinted from Harary (1959).

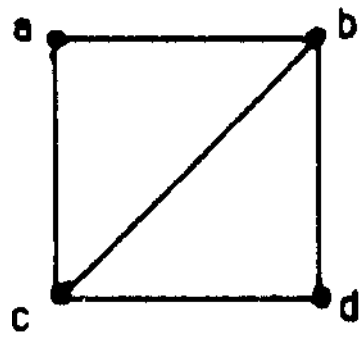
Figures on the right--reprinted form French, J. (1956) A formal theory of social power. Psychological Review, 63, 181-194.

Figure 7. Cluster Analyses of the AIDS data using UCLUS and the COMPLETE LINK algorithm.

Figure 8. Graphical representation of the AIDS network showing the structurally and stochastically equivalent nodes within clusters.

Figure 1

1a.



1b.

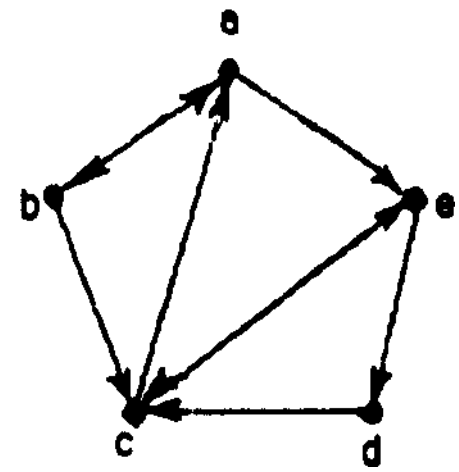
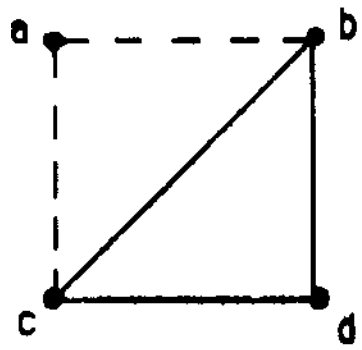


Figure 2

2a.



2b.

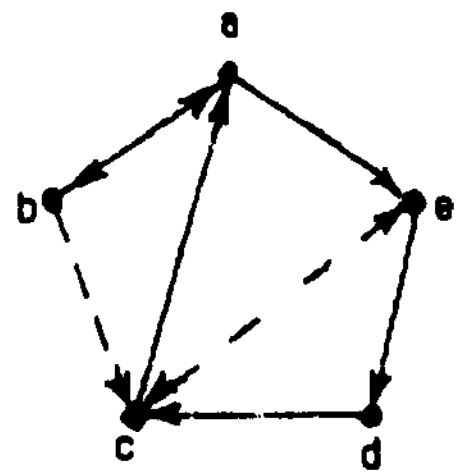


Figure 3

3a.

	a	b	c	d
a	0	1	1	0
b	1	0	1	1
c	1	1	0	1
d	0	1	1	0

3b.

	a	b	c	d	e
a	0	1	0	0	1
b	1	0	1	0	0
c	1	0	0	0	1
d	0	0	1	0	0
e	0	0	1	1	0

Figure 4

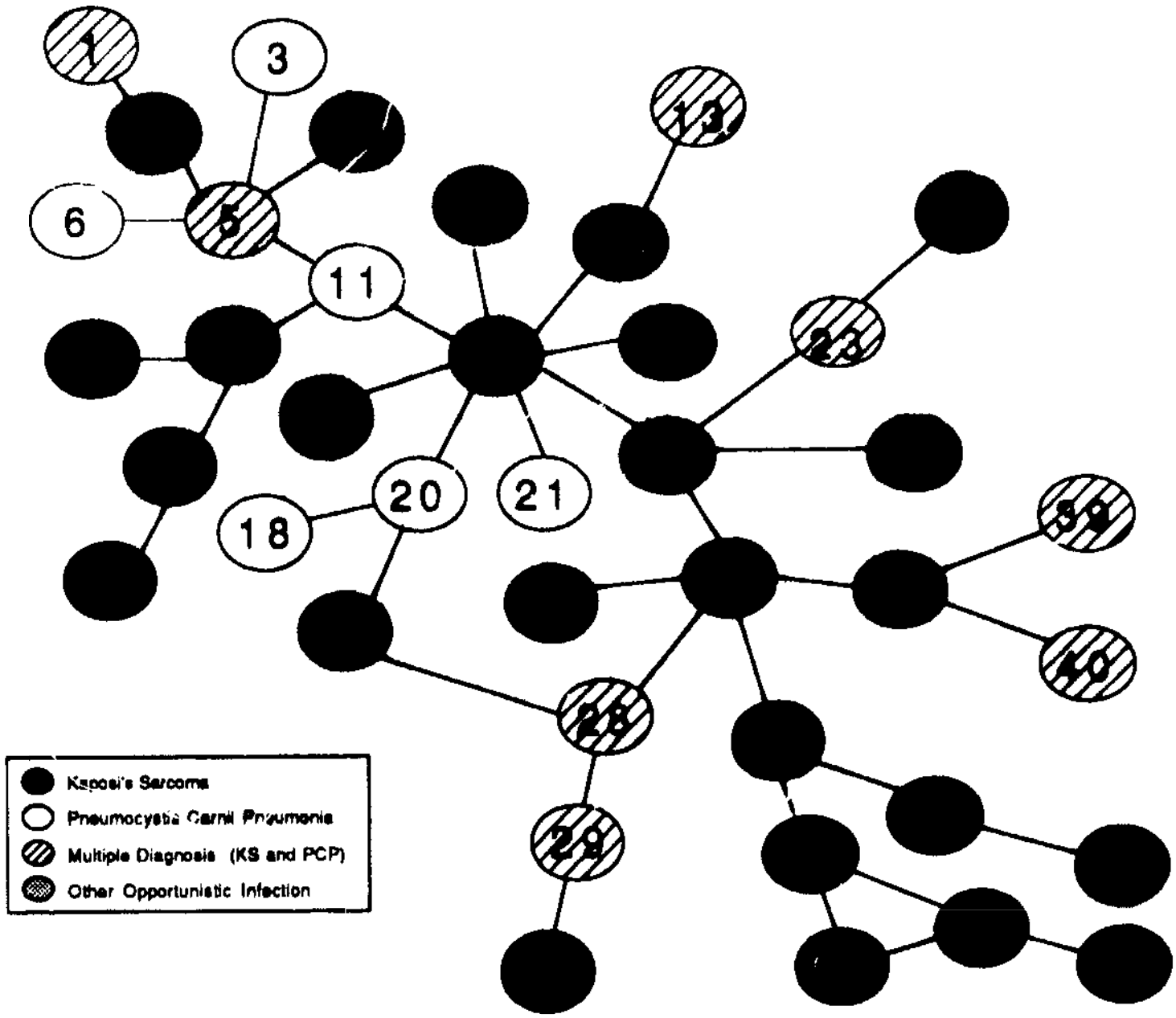
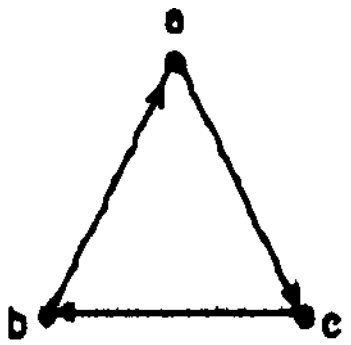

1
3
13
6
5
11
23
20
21
18
39
40
28
29
Kaposi's Sarcoma
Pneumocystis Carnii Pneumonia
Multiple Diagnosis (KS and PCP)
Other Opportunistic Infection

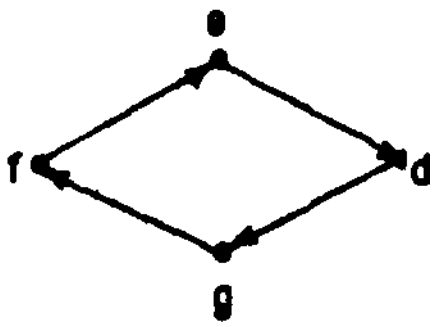
Figure 5

i, j =	
1	0100000000000000000000000000000000000000
2	1000100000000000000000000000000000000000
3	0000100000000000000000000000000000000000
4	0000100000000000000000000000000000000000
5	0111010000100000000000000000000000000000
6	0000100000000000000000000000000000000000
7	0000000100000000000000000000000000000000
8	0000001000000000000000000000000000000000
9	0000000101000000000000000000000000000000
10	0000000010000000000000000000000000000000
11	0000100100000001000000000000000000000000
12	0000000000000001000000000000000000000000
13	0000000000000100000000000000000000000000
14	0000000000001001000000000000000000000000
15	0000000000000010000000000000000000000000
16	0000000000110110100111000000000000000000
17	0000000000000001000000000000000000000000
18	0000000000000000100000000000000000000000
19	0000000000000000100000001000000000000000
20	0000000000000010110000000000000000000000
21	0000000000000010000000000000000000000000
22	0000000000000010000001011000000000000000
23	0000000000000000000101000000000000000000
24	0000000000000000000010000000000000000000
25	0000000000000000000010000000000000000000
26	0000000000000000000010000110010000001000
27	0000000000000000000000100000000000000000
28	0000000000000000010000001001000000000000
29	0000000000000000000000000010100000000000
30	0000000000000000000000000100000000000000
31	0000000000000000000000001000001000010000
32	0000000000000000000000000000001011000000
33	0000000000000000000000000000001010000000
34	0000000000000000000000000000001101000000
35	0000000000000000000000000000000010000000
36	0000000000000000000000000000001000001000
37	0000000000000000000000000000000010000000
38	0000000000000000000000000010000000000011
39	0000000000000000000000000000000000000100
40	0000000000000000000000000000000000000100

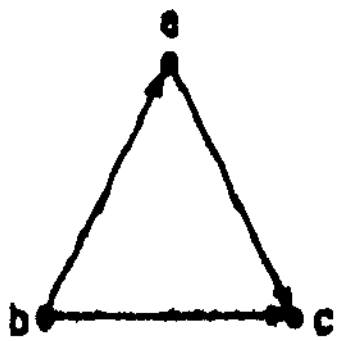
Figure 6

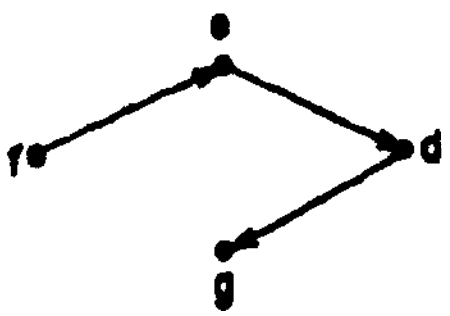
strong



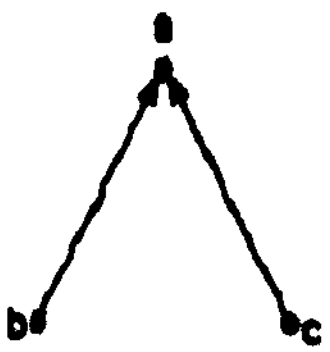


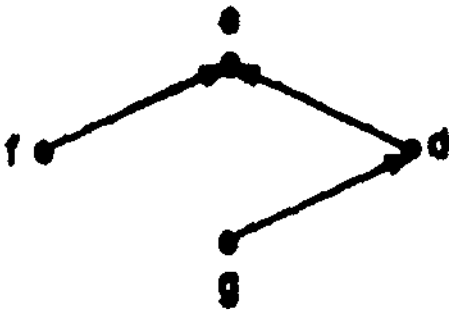
unilateral



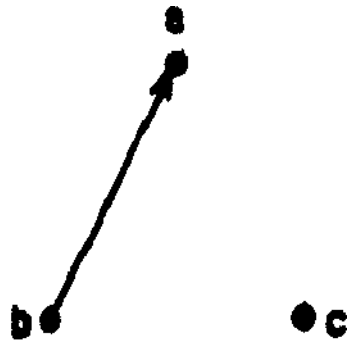


week





disconnected



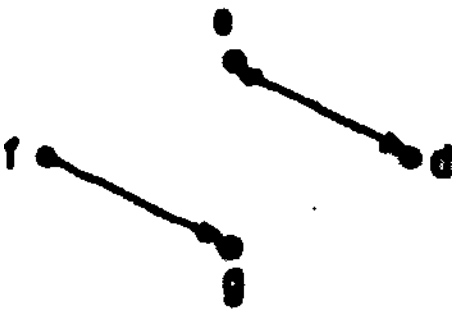
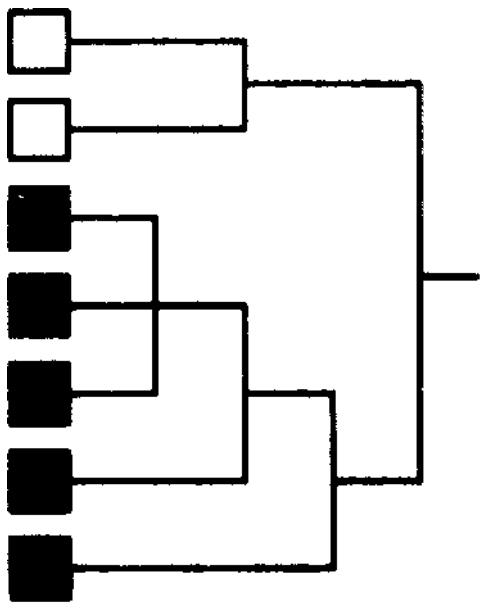


Figure 7

UCLUS



Nodes

1 2 3 4 5 6

7 8 9 10

12 13 14 15 16 17 21 11

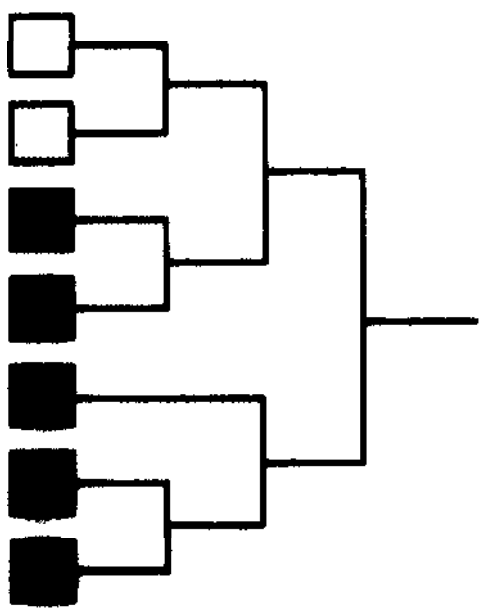
22 23 24 25

18 20 19 28

26 27 38 39 40 29 30

32 33 34 36 31 36 37

COMPLETE LINK



Nodes

1 2 3 4 5 6

7 8 9 10 11

12 13 14 15 16 17 21 18 20

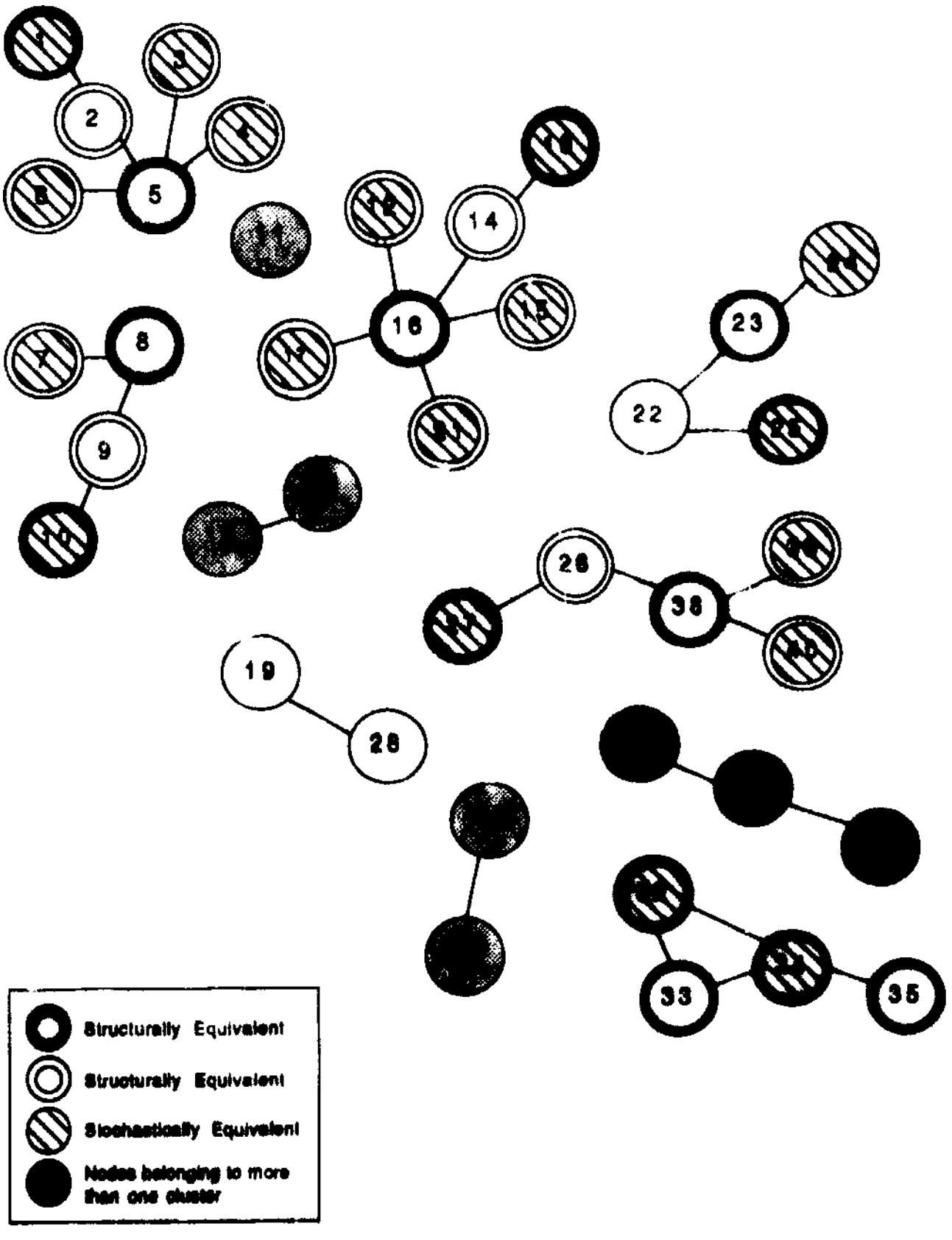
22 23 24 25

19 28 29 30

26 27 38 39 40 31 36 37

32 33 34 35

Figure 8


2
5
14
16
8
9
23
22
28
38
19
28
33
35
Structurally Equivalent
Structurally Equivalent
Stochastically Equivalent
Nodes belonging to more than one cluster